****

**Impact of Goods and Service Tax in India (GST)**

Submitted by: -

Surendra Kumar Dattatrey – 71710015

Table of Contents

[1. Abstract 3](#_Toc484185353)

[2. Business problem 3](#_Toc484185354)

[3. Introduction 3](#_Toc484185355)

[3.1 Background 3](#_Toc484185356)

[3.2 Purpose / Motivation 3](#_Toc484185357)

[3.3 Audience 3](#_Toc484185358)

[4. Significance of the study/Learning objective 4](#_Toc484185359)

[5. Research questions and hypothesis 4](#_Toc484185360)

[6. Methodology and Tools 4](#_Toc484185361)

[7. Data 4](#_Toc484185362)

[7.1 Data collection 4](#_Toc484185363)

[7.2 Data cleaning 5](#_Toc484185364)

[8. Analysis and Results 5](#_Toc484185365)

[9. Conclusion and recommendations 10](#_Toc484185366)

[10. Limitations 10](#_Toc484185367)

[11. References 10](#_Toc484185368)

[12. Appendix 11](#_Toc484185369)

# 1. Abstract

**Goods and Services Tax (GST)** is a comprehensive [indirect tax](https://en.wikipedia.org/wiki/Indirect_tax) on manufacturing, sales and consumption of goods and services throughout in India to replace old traditional taxes levied by the [central](https://en.wikipedia.org/wiki/Govt_of_India) and [state](https://en.wikipedia.org/wiki/States_of_India) governments. Finally Bill on The Goods and Service Tax (GST) is passed by Union Government of India. It is the biggest tax reform to boost Indian economy. This tax reform is impacting almost all sectors in India. So different stakeholders are reacting differently due to the challenges they are facing due to implementation of GST.

Implementation of GST would have an impact on different industries; resultant different sectors are reacting on this reform differently.

This paper is on sentiment analysis on overall GST along with four core sectors Exports, Automobile, and Manufacturing. To perform analysis 60 articles extracted from internet on each topic.

# 2. Business problem

By this study, I tried to find out sentiments on over all GST and specific to few industries. To check is there any difference between sentiments. Also try to check that sentiments across the industries are positive, negative or Neutral.

# 3. Introduction

## 3.1 Background

I carried out a study by which I tried to check sentiments on GST and specific to industries. This sentiment analysis is on the basis of existing libraries which are being used in R.

## 3.2 Purpose / Motivation

By this study, I will get an understanding how industry is thinking about GST. Does it much better compare to old traditional tax system or creating more problems.

Basically I am trying to find out an answer of below question:

*What are people thinking about GST? Does industries had positive sentiment, Negative sentiments or Neutral sentiments?*

## 3.3 Audience

This study would be useful to Indian government, Mango people of India.

# 4. Significance of the study/Learning objective

In this study, I will get an understanding on sentiments of the World Wide Web and sentiments from Industries.

Learning objective for this study to understand the Sentiment analysis and how this can be used on any business context

# 5. Research questions and hypothesis

With this study I am trying to research the impact of GST on Industries. I would like figure out the answers of following questions:

1. Does Industries thinks positively or negatively on GST? If they thinks negatively what are

2. What World Wide Web thinks about GST?

3. Does stock price changes when investor is more attentive?

# 6. Methodology and Tools

This paper is being presents on sentimental analysis of impact of GST on three sectors Automobile, Export and Manufacturing.

* + - Web crawl the data from 60 web sites for each industry sector (which are given above).
    - On the basis of extracted data, I tried to check for positive and negative Tokens / Terms used in articles.
    - I also plotted positive terms against the negative terms on a graph, to calculated overall polarity of the articles**.**
    - Analyze the main topics discussed in the articles for a particular sector

I used Python for web crawling of data from Google and by using R, I performed the analysis.

# 7. Data

## 7.1 Data collection

Google was predominantly used as secondary data source for this study.

Used following keywords for extracting data:

1. GST impact – For general sentiment analysis
2. GST impact on Exports
3. GST impact on Automobile
4. GST impact on Manufacturing

I used above keyword to search data from “www.google.co.in”, downloaded data from top 60 web sites and excluding all .pdf file and cleaned the corpus (removing regular expressions and stop words like – the , an, a, again etc. These stop words are not required in the analysis.

## 7.2 Data cleaning

Extracted corpus for all the keywords and performed multiple cleaning process on the extracted data:

1. Removed all Regular expressions
2. Removing all stop words from above extracted corpus

# 8. Analysis and Results

For the impact of GST, global concern “GST impact” along with three core industry segments Automobile, Export, and Manufacturing has been analyzed.

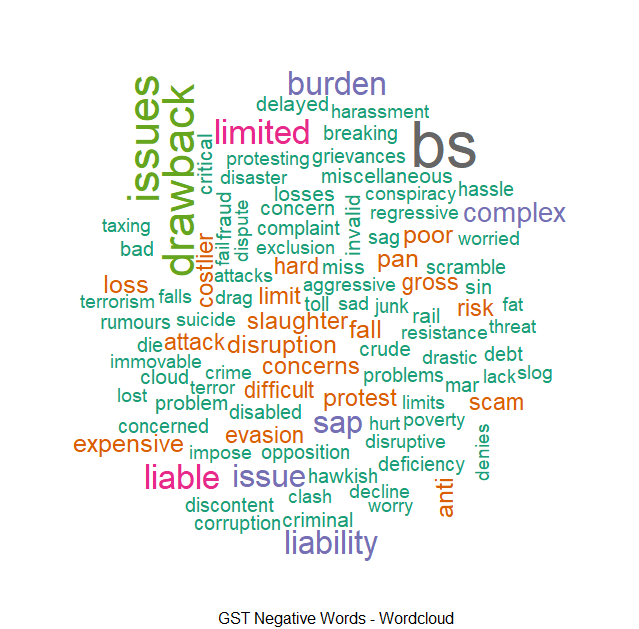
GST Impact seems to be quite positive with some negatives. On positive side where people are talking about “Benefits”, “Reform”, “Ready” etc. and on negative side people are talking about “issues”, “drawbacks”, “liability” etc.

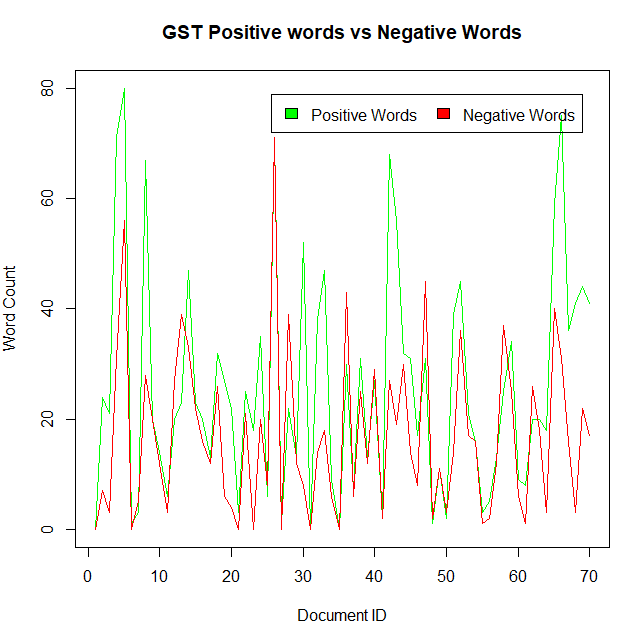
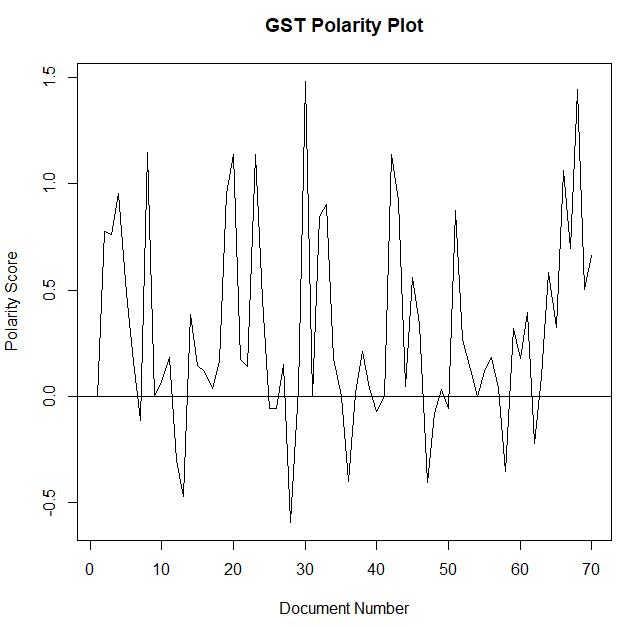
**8.1 GST impact –**

For over all “GST impact”, sentiments are on more positive side. On positive side, people are seeing benefits, easy process, reforms, and faster things, on the other hand on negative side people are thinking about drawbacks, liabilities, complex process etc.

To get more details on this, I tried to plot “Word Cloud Chart”, named as “GST Positive Words v/s Negative Word”, in that chart you can see that frequency of the positive word is much higher than negative words.

I also draw a polarity chart to check polarity of positive or negative words (which tells you which sentiment had more weightage), in that graph you can see that positive word polarity is very high compare to negative word polarity.

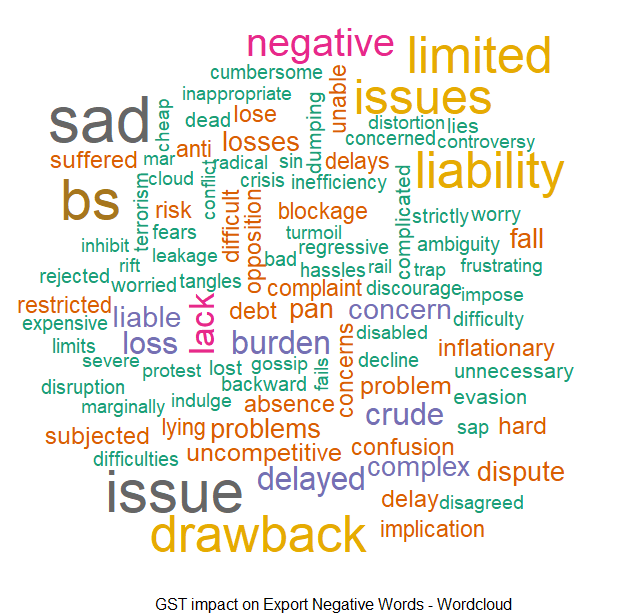
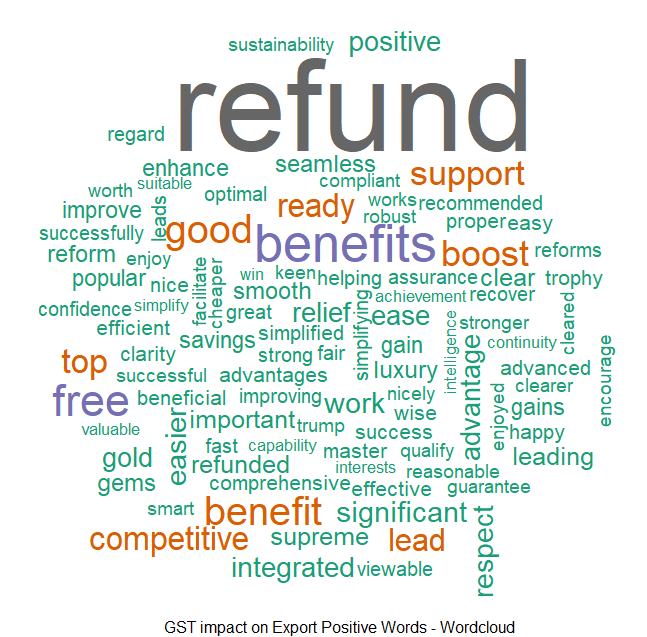
 

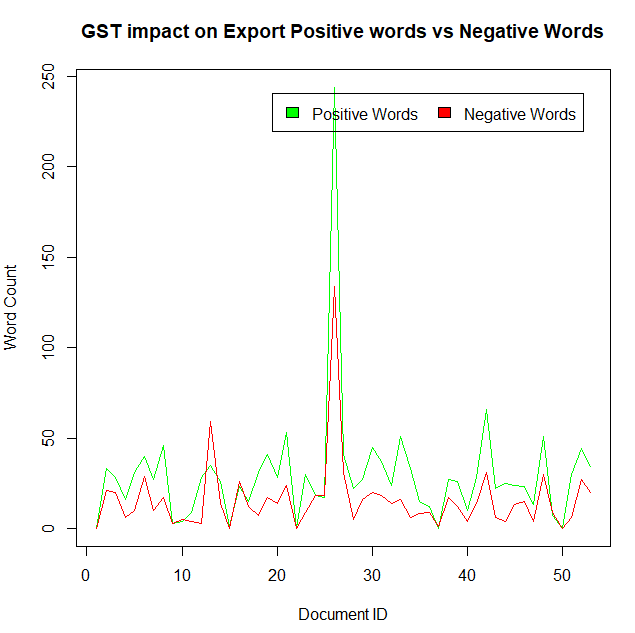
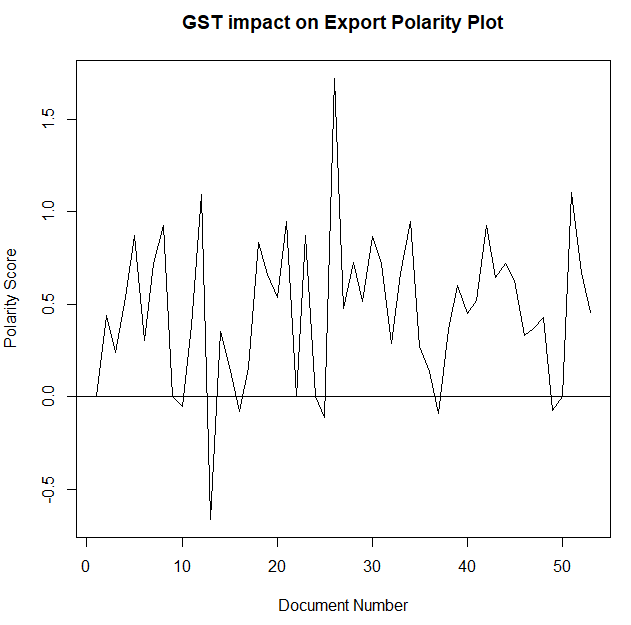
**8.2** **GST Impact on Export**

For over all “GST impact on Export”, sentiments are on positive side. People can expect to save significantly (with Refunds), it would boost the export, people can live luxury life, and there would be more benefits. This will in turn increase the demand and hence increase the markets for Export players, on the other hand on negative side people are thinking about drawbacks, issue etc., may be they are talking about some issue with GST process (which they are feeling) and talking about complex process with GST etc.

To get more details on this, I tried to plot “Word Cloud Chart”, named as “GST impact on Export Positive Words v/s Negative Word”, in that chart you can see that frequency of the positive word is much higher than negative words.

I also draw a polarity chart to check polarity of positive or negative words (which tells you which sentiment had more weightage), in that graph you can see that positive word polarity is very high compare to negative word polarity.



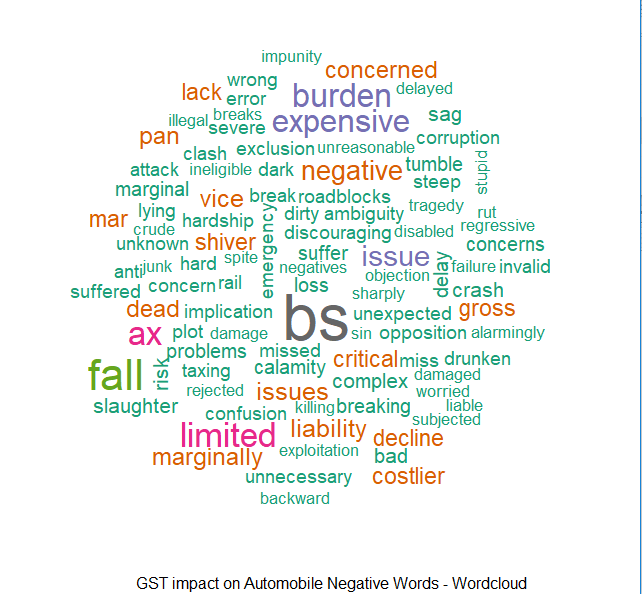
 

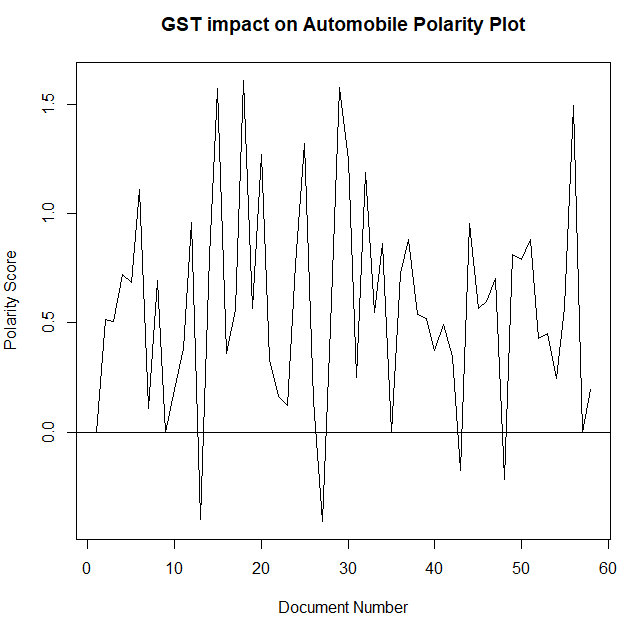
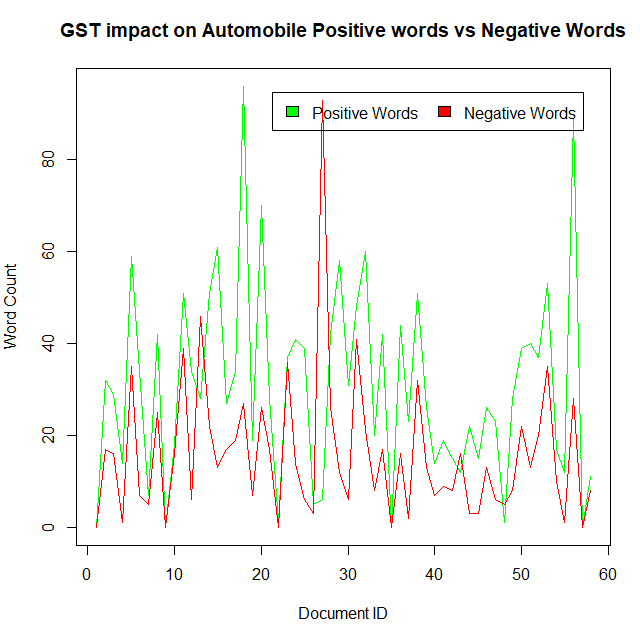
**8.3 GST Impact on Automobile**

For over all “GST impact on Automobile”, sentiments are mixed of feeling. People can see benefits, people can see that thing would be cheaper, people can live luxury life, and there would be more benefits also many people thinks that Automobile would be expensive, additional burden on customers, sales could fall, also people are concerned about it.

To get more details on this, I tried to plot “Word Cloud Chart”, named as “GST impact on Automobile Positive Words v/s Negative Word”, in that chart you can see that frequency of the positive word and negative words are almost equal.

I also draw a polarity chart to check polarity of positive or negative words (which tells you which sentiment had more weightage), in that word cloud graph, overall sentiments are positive for automobiles sector but there are some articles which had negative polarity as well.

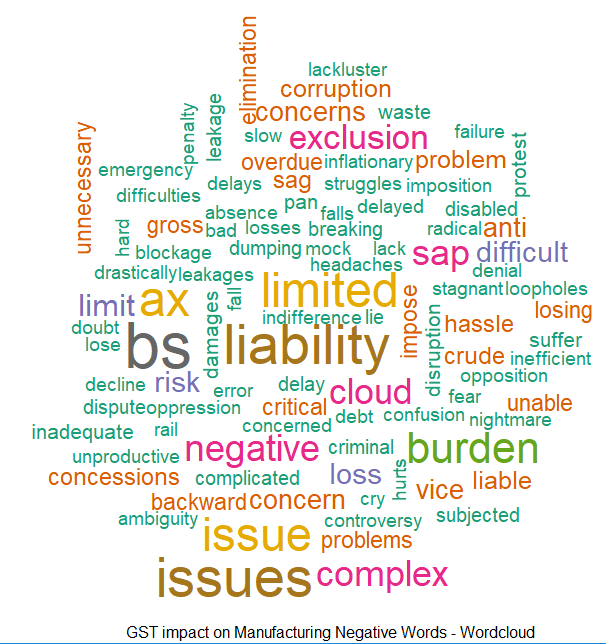


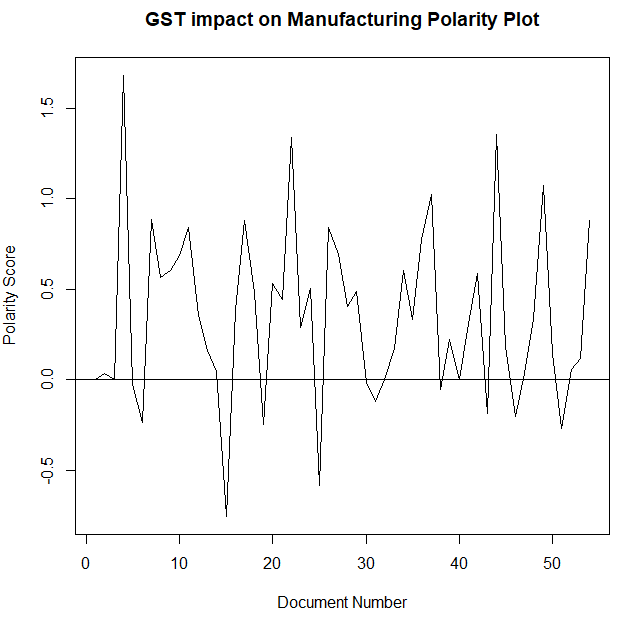
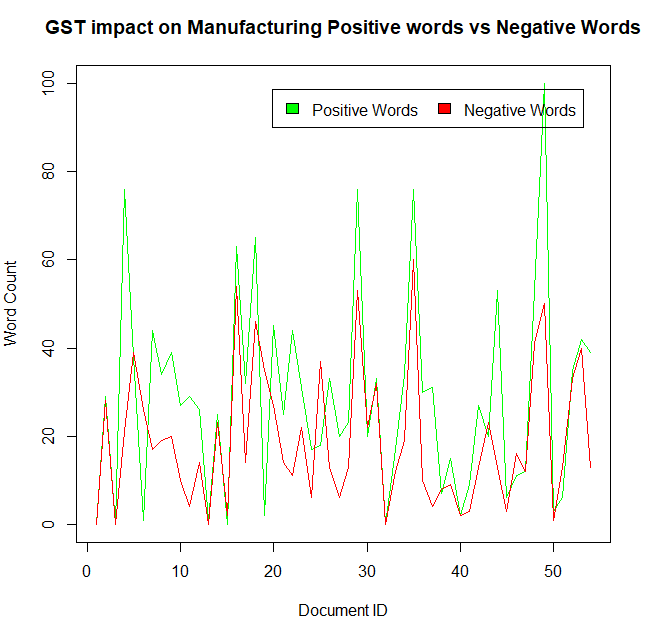
**8.4 GST Impact on Manufacturing**

For over all “GST impact on Manufacturing”, sentiments are mixed of feeling. People can see benefits, importance, also can see that manufacturing industry would be improve, things would be smooth. Etc, also many people thinks that there would be issues in GST, should provide concessions to industry, it would be disruption to the industry etc.

To get more details on this, I tried to plot “Word Cloud Chart”, named as “GST impact on Manufacturing Positive Words v/s Negative Word”, in that chart you can see that frequency of the positive word and negative words are almost equal (more on positive side).

I also draw a polarity chart to check polarity of positive or negative words (which tells you which sentiment had more weightage), in that word cloud graph, overall sentiments are positive for automobiles sector but there are some articles which had negative polarity as well.



# 9. Conclusion and recommendations

All the Three sectors Automobile, Export and Manufacturing had over all positive sentiments in implementation of GST.

For Export, it had more positive where people see themselves affording luxury products. They are expecting decrease in prices.

For Automobile, there are mixed feeling, where people are more clarity on new tax scheme. They see more benefits and cheaper automobiles. On the other hand, people feeling that automobile would be costlier, items would be expensive and burden on customer post implementation of GST.

For Manufacturing, there are mixed feeling. They see more benefits, important and ready for GST. On the other hand, people are thinking that there would be issues in GST, would be disruption to industry etc.

# 10. Limitations

This paper has some limitations. Sentiment analysis was done on limited topics. Only 60web links used to perform this analysis. Also analysis towards social media like twitter and Facebook also not covered in this analysis.

This analysis is giving us some context about GST but not fully. After adding data from social media, this study can be done in much broader way. Apart from those only 3 industries along with over all GST is covered in this study. We can add all other areas which are impacted due to GST in study.

# 11. References

Searched for following topics on <https://www.google.co.in> and scraped data form first 60 web articles.

1. “impact of GST on Automobile”
2. “impact of GST on Exports”
3. “impact of GST on Manufacturing”
4. “impact of GST”

# 12. Appendix

R Code which I used for sentiment analysis. In this code, I connected to respective file name for that particular industry on which I was trying to do sentiment analysis.

rm(list=ls()) # Clear workspace

#--------------------------------------------------------#

# Step 0 - Assign Library & define functions #

#--------------------------------------------------------#

library(text2vec)

library(data.table)

library(stringr)

library(tm)

library(RWeka)

library(tokenizers)

library(slam)

library(wordcloud)

library(ggplot2)

text.clean = function(x) # text data

{ require("tm")

x = gsub("<.\*?>", " ", x) # regex for removing HTML tags

x = iconv(x, "latin1", "ASCII", sub="") # Keep only ASCII characters

x = gsub("[^[:alnum:]]", " ", x) # keep only alpha numeric

x = tolower(x) # convert to lower case characters

x = removeNumbers(x) # removing numbers

x = stripWhitespace(x) # removing white space

x = gsub("^\\s+|\\s+$", "", x) # remove leading and trailing white space

return(x)

}

#--------------------------------------------------------#

# Step 1 - Reading text data #

#--------------------------------------------------------#

# temp.text = readLines('C:\\Users\\30773\\Desktop\\Data Science\\cba batch 7\\text classification\\testdata.txt')

temp.text = readLines(file.choose()) # iron man reviews

# head(temp.text,1)

data = data.frame(id = 1:length(temp.text), text = temp.text, stringsAsFactors = F)

# dim(data)

# Read Stopwords list

stpw1 = readLines('C:\\Users\\Yash\\Desktop\\ISB Documents\\Term 1\\TABA\\Amit Session Files\\Session 1 Materials-20170424\\stopwords.txt') #

stpw2 = tm::stopwords('english') # tm package stop word list; tokenizer package has the same name function

comn = unique(c(stpw1, stpw2)) # Union of two list

stopwords = unique(gsub("'"," ",comn)) # final stop word lsit after removing punctuation

x = text.clean(data$text) # pre-process text corpus

x = removeWords(x,stopwords) # removing stopwords created above

x = stripWhitespace(x) # removing white space

# x = stemDocument(x)

#--------------------------------------------------------#

###### Create DTM using text2vec package #

#--------------------------------------------------------#

t1 = Sys.time()

tok\_fun = word\_tokenizer

it\_m = itoken(x,

# preprocessor = text.clean,

tokenizer = tok\_fun,

ids = data$id,

progressbar = T)

vocab = create\_vocabulary(it\_m

# ngram = c(2L, 2L),

#stopwords = stopwords

)

pruned\_vocab = prune\_vocabulary(vocab,

term\_count\_min = 1)

# doc\_proportion\_max = 0.5,

# doc\_proportion\_min = 0.001)

vectorizer = vocab\_vectorizer(pruned\_vocab)

dtm\_m = create\_dtm(it\_m, vectorizer)

dim(dtm\_m)

dtm = as.DocumentTermMatrix(dtm\_m, weighting = weightTf)

a0 = (apply(dtm, 1, sum) > 0) # build vector to identify non-empty docs

dtm = dtm[a0,] # drop empty docs

print(difftime(Sys.time(), t1, units = 'sec'))

#--------------------------------------------------------#

# Sentiment Analysis #

#--------------------------------------------------------#

require(qdap) || install.packages("qdap") # ensure java is up to date!

library(qdap)

x1 = x[a0] # remove empty docs from corpus

t1 = Sys.time() # set timer

pol = polarity(x1) # Calculate the polarity from qdap dictionary

wc = pol$all[,2] # Word Count in each doc

val = pol$all[,3] # average polarity score

p = pol$all[,4] # Positive words info

n = pol$all[,5] # Negative Words info

Sys.time() - t1 # how much time did the above take?

head(pol$all)

head(pol$group)

positive\_words = unique(setdiff(unlist(p),"-")) # Positive words list

negative\_words = unique(setdiff(unlist(n),"-")) # Negative words list

print(positive\_words) # Print all the positive words found in the corpus

print(negative\_words) # Print all neg words

#--------------------------------------------------------#

# Create Postive Words wordcloud #

#--------------------------------------------------------#

pos.tdm = dtm[,which(colnames(dtm) %in% positive\_words)]

m = as.matrix(pos.tdm)

v = sort(colSums(m), decreasing = TRUE)

windows() # opens new image window

wordcloud(names(v), v, scale=c(8,1),1, max.words=100,colors=brewer.pal(8, "Dark2"))

title(sub = "GST impact on Manufacturing Positive Words - Wordcloud")

# plot barchart for top tokens

test = as.data.frame(v[1:15])

windows() # opens new image window

ggplot(test, aes(x = rownames(test), y = test)) +

geom\_bar(stat = "identity", fill = "blue") +

geom\_text(aes(label = test), vjust= -0.20) +

theme(axis.text.x = element\_text(angle = 90, hjust = 1))

#--------------------------------------------------------#

# Create Negative Words wordcloud #

#--------------------------------------------------------#

neg.tdm = dtm[,which(colnames(dtm) %in% negative\_words) ]

m = as.matrix(neg.tdm)

v = sort(colSums(m), decreasing = TRUE)

windows()

wordcloud(names(v), v, scale=c(4,1),1, max.words=100,colors=brewer.pal(8, "Dark2"))

title(sub = "GST impact on Manufacturing Negative Words - Wordcloud")

# plot barchart for top tokens

test = as.data.frame(v[1:15])

windows()

ggplot(test, aes(x = rownames(test), y = test)) +

geom\_bar(stat = "identity", fill = "red") +

geom\_text(aes(label = test), vjust= -0.20) +

theme(axis.text.x = element\_text(angle = 90, hjust = 1))

#--------------------------------------------------------#

# Positive words vs Negative Words plot #

#--------------------------------------------------------#

len = function(x){

if ( x == "-" && length(x) == 1) {return (0)}

else {return(length(unlist(x)))}

}

pcount = unlist(lapply(p, len))

ncount = unlist(lapply(n, len))

doc\_id = seq(1:length(wc))

windows()

plot(doc\_id,pcount,type="l",col="green",xlab = "Document ID", ylab= "Word Count")

lines(doc\_id,ncount,type= "l", col="red")

title(main = "GST impact on Manufacturing Positive words vs Negative Words" )

legend("topright", inset=.05, c("Positive Words","Negative Words"), fill=c("green","red"), horiz=TRUE)

# Documet Sentiment Running plot

windows()

plot(pol$all$polarity, type = "l", ylab = "Polarity Score",xlab = "Document Number")

abline(h=0)

title(main = "GST impact on Manufacturing Polarity Plot" )